

# Report

By:- Mayank Singh Kushwaha  
2017CS50413

In my initial naïve bayes model without feature engineering, we get the accuracy of 83.94.

When stop words were removed and stemming was implemented, I got an accuracy of 84.17.

For feature engineering, I just considered bigram, along with removing stop words. Due to less time available I didn't have time to implement any other feature engineering.

I included bigrams along with unigram. I also replaced dots, question mark and comma with space and stop words were not included in bigram. This increased the accuracy to 86.285(best accuracy).

Summary:

Accuracy with no Stopping and Stemming = 83.94

Accuracy with Stopping and Stemming = 84.17

Accuracy with Stopping and bi-grams = 86.285